Why empirical cost functions get scale economies wrong

Lijesen, M.G.

2009

document version
Early version, also known as pre-print

Why empirical cost functions get scale economies wrong

Mark G. Lijesen
mlijesen@feweb.vu.nl
VU University Amsterdam
Department of Spatial Economics
De Boelelaan 1105
1081 HV Amsterdam
tel +31205988487

Abstract
Empirical applications of the translog cost function often conclude that firms operate at increasing returns to scale. From the viewpoint of economic theory, this does not make sense. We demonstrate that empirical cost functions ignore the fact that differences in firm output depends on cost differences between firms. We show graphically and mathematically that ignoring this mechanism leads to an overestimation of returns to scale. We propose a slightly altered specification and test it empirically. The empirical results show that the alternative specification yields better statistical results and is consistent with economic theory.

keywords: costs, firm size, translog, economies of scale

JEL-codes: L11, L25, L93

1 introduction

The empirical analysis of firms' cost (and production) functions plays an important role in industrial economics. Cost functions provide useful information on the cost parameters and cost structure of an industry. Increasing computer power and data availability have induced a large number of empirical papers, often combined with stochastic frontiers. (see e.g Pels et al., 2003; Fries and Taci, 2005; Goh and Yong, 2006) One of the main issues in empirical cost and production functions is the assessment of returns to scale in production. Empirical knowledge of returns to scale may play an important role in assessing the regulator's response to industry consolidation (see e.g. Farrel and Shapiro, 2000).

Many empirical studies in various industries find returns to scale to be increasing or not significantly different from constant, as Table 1 suggests. This could be true for some sectors and some specifications of the cost function. In general however, increasing returns to scale are inconsistent with economic theory, and even constant returns to scale are not very likely. The influential paper on entry by Bresnahan and Reiss (1991) hinges on the condition that marginal costs are increasing for any market with more than one firm. In fact, it is pretty straightforward that a competitive equilibrium with fixed costs requires decreasing returns to scale, whereas increasing returns to scale are limited to (natural) monopoly markets (see e.g. Panzar and Willig, 1977).

The discrepancy between the empirical studies and economic theory stems from the fact that the empirical cost functions implicitly take quantities as exogenous, whereas theory states that efficient firms grab a larger market share. In this paper, we will show that taking into the endogenous nature of quantities will lead to a different economic interpretation of empirical cost functions, and often to a different conclusion with respect to returns to scale.

The remainder of this paper is organized as follows. Section 2 takes a brief look into recent empirical applications of the translog cost function and explores the conclusions. In section 3, we develop a simple model linking scale economies and market equilibrium and explaining why empirical cost functions fail to measure scale economies correctly. Section 4 follows up on the model, exploring what empirical cost functions do tell us about scale economies and how empirical results can be interpreted. The data and empirical illustration of our findings are then described in section 5. Section 6 concludes the paper.

2. Findings in the recent literature

A fairly large number of studies using the translog cost function has been published recently. We limit ourselves to studies that have appeared over the last fifteen years and that apply to markets where entry and exit are to a certain extent free. This leaves out studies regarding utilities, railways and so on. The majority of the studies focus on banks, mainly because of the availability of good quality data and the policy relevance of scale economies with respect to mergers. Table 1 summarizes the results from these studies

Table 1 *Summary of findings in recent empirical translog cost functions*

reference	remark	industry	Scale economies[a]	sign of parameter of quadratic term[b]
Altunbaș *et al.*, 2001		banks	IRS	**positive**
Altunbaș *et al.*, 1997		banks	DRS	**positive**
Berger *et al.*, 1997		banks	IRS	-*
Chua *et al.*, 2005	concavity imposed	airlines	IRS	positive
	not imposed		**IRS**	negative
Fries and Taci, 2005		banks	CRS	**positive**
Goh and Yong, 2006	model A	airlines	IRS	negative
	model B		IRS	negative
	model C		IRS	**negative**
Hansen and Zwanziger, 1996	California 1981	hospitals	CRS	mixed
	California 1985		CRS	mixed
	New York 1981		CRS	mixed
	New York 1985		CRS	mixed
	Canada 1981		CRS	mixed
	Canada 1985		CRS	mixed
Hughes and Mester, 1993		banks	IRS	mixed
Hughes and Mester, 1998		banks	**IRS**	-*
Humphrey, 1993		banks	DRS	**positive**
Malhotra *et al.*, 2001	1999 data	retail superannuation funds	CRS	negative
	2000 data		CRS	positive
McCarthy and Urmanbetova, 2006		pulp and paper industry	IRS	negative
Mountain and Thomas, 1999		banks	DRS	positive

[a]: bold: scale elasticity significantly different from one

[b]: bold: parameter significantly different from zero

*: not reported

The fourth column in table 1 shows that increasing returns to scale are often found in the empirical literature, and for many different industries. We found only three studies where economies of scale are significantly different from unity (implying CRS). It should be noted though that many studies do not report the statistical significance of the scale parameter. The right hand column provides information on the shape of the cost function. The translog cost function is famous for its very flexible functional form, but flexibility comes at a cost. Concavity is not guaranteed, whereas one would expect a cost function to be concave. A positive sign for the parameter of the quadratic term implies that the cost function is concave, which is not always the case for all goods.

Some of the papers mentioned in the table are merely good and relevant applications of translog cost functions, whereas some of the others also address methodological issues. Chue *et al.* (2005) discuss the issue of concavity of the cost function, which is somewhat related to the issue addressed in our paper. Chue et al. (2005) impose local concavity on a translog cost function of airlines and note that this leads to a shift in conclusions. Not only does the sign of the parameter of the quadratic term switch, but the conclusion on scale economies also changes. The unrestricted version of the model suggests significant economies of scale, whereas the version with imposed concavity severely weakens this conclusion. This finding is much in line with ours, as we will see in the following section. Hughes and Mester (1993, 1998) find increasing returns to scale for banks. They conclude that scale economies are to a large extent caused by the fact that larger banks are perceived safer and therefore pay lower risk premiums.

3. Scale economies and market equilibrium

In this section, we develop a simple model describing firm's behaviour in a market with free entry and fixed costs. Firms maximize profits:

$$\pi = PX - C \tag{1}$$

The backbone of the model is a single product translog cost function. Input prices are omitted for analytical tractability:

$$lnC = F + \gamma lnX + \frac{1}{2}\theta lnX^2 \tag{2}$$

Where *F* denotes fixed costs and $\gamma + \theta lnX$ represents the cost elasticity of X. If $\gamma + \theta lnX < 1$, the cost function exhibits increasing returns to scale, $\gamma + \theta lnX = 1$ denotes constant returns to scale and $\gamma + \theta lnX > 1$ implies the presence of decreasing returns to scale. From the log differential, we may define marginal costs as $(\gamma + \theta lnX)\frac{C}{X}$. In the competitive equilibrium, price equals marginal costs:

$$P = (\gamma + \vartheta lnX)\frac{C}{X} \tag{3}$$

Substitution of (2) and (3) into the profit function (1) yields a profit of:

$$\pi = (\gamma + \theta lnX - 1)C \tag{4}$$

For profits to be positive, we require:

$$\gamma + \theta lnX \geq 1 \tag{5}$$

This implies that constant or decreasing returns to scale at the firm level are a necessary condition for profitability. This is simply a slightly different way of formulating the straightforward textbook knowledge that "…the supply curve of a competitive firm must lie along the upward sloping part of the marginal cost curve." (Varian, 2006, p. 388).

In imperfect markets, this conclusion does not hold per se, but it is fairly straightforward to argue that increasing returns to scale will lead to cut-throat price competition and hence the exit of firms out of the market until constant (or decreasing) returns to scale are reached or only one firm is left. In equilibrium, increasing returns to scale is therefore feasible for natural monopoly markets only, whereas constant returns to scale in the presence of fixed costs can only prevail if firms have sufficient market power.

Based on our analysis above we conclude that the vast majority of markets should operate at decreasing returns to scale. This is however not what we observe from empirical studies. We will argue below that the discrepancy between theory and practice follows from the implicit assumption in empirical cost studies that the firm-level production does not depend on costs.[1] We will adjust the model outlined so far by allowing for cost differences between firms and assess how this affects their individual output levels. We adjust cost equation (2) to account for differences in cost levels, allowing parameter *γ* to vary between firms:

$$lnC_i = F + \gamma_i lnX_i + \frac{1}{2}\theta lnX_i^2 \tag{6}$$

Firms equate marginal costs to marginal revenue to maximize profits.:

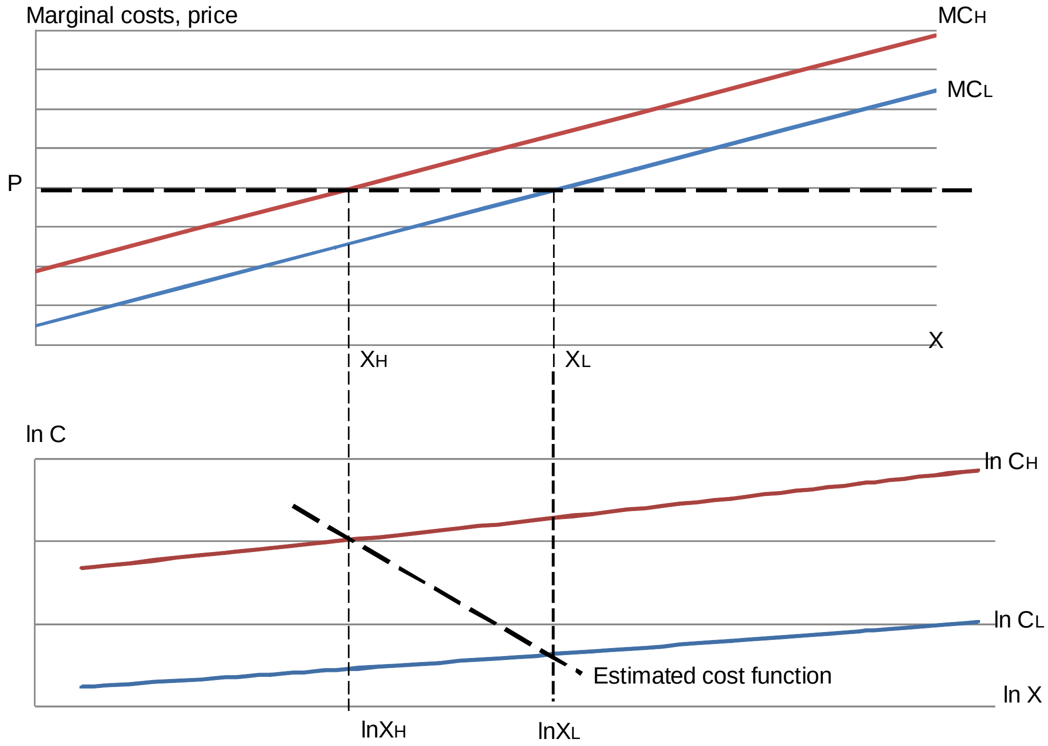
$$P(X_i) + X_i\frac{\delta P}{\delta X_i} = (\gamma_i + \theta lnX_i)\frac{C_i}{X_i} \tag{7}$$

Where $\frac{\delta P}{\delta X_i} \leq 0$ can be linked directly to the conjectural variations theory. (see e.g. Bresnahan, 1981 and Perry, 1982). The conjectural variation model assumes that each firm expects (conjectures) an output reaction by rival firms as a function of its own output. In a simple two firm game, the conjectural variations parameter may be written as $v_i = \frac{\delta X_j}{\delta X_i}$. The value of v_i can ten then be interpreted as a measure for the ferocity of competition. If $v_i = -1$, firms behave consistent with the Bertrand model, whereas a market with $v_i = 0$ can be best described by the Cournot model. Monopoly or cartel markets are characterized by a unity value for the conjectural variations parameter. Sticking to the simple two-firm case for instructional purposes and using a simple inverse demand relationship, such as $P = \alpha - \beta(X_i + X_j)$, we note that $\frac{\delta P}{\delta X_i} = -\beta(1 + v_i)$, which equals zero for the Bertrand or competitive equilibrium.

[1] Torres and Morrison Paul (2006) also address the issue of output endogeneity, but they do so from the angle of demand uncertainty.

Equation (7) reveals that marginal costs are equal for all firms in the competitive equilibrium. Figure 1 below illustrates that cost differences in a competitive market may lead to erroneous conclusions regarding economies of scale. The upper half of figure 1 shows two firms with different marginal cost levels and otherwise identical cost functions.[2] It is immediately clear that both firms operate at decreasing returns to scale, as the marginal cost curve is upward sloping. Since both firms are price takers, the low cost firm produces more than the high cost firm (X_L>X_H). The lower half of the figure shows how this translates to what the empirical researcher observes. She observes only two points: (lnX_H, lnC_H) and (lnX_L, lnC_L) in the lower half of figure 1. We can hardly blame the empirical researcher for concluding that the cost structure in this market inhibits increasing returns to scale, whereas all firms are in fact experiencing decreasing returns to scale.

Figure 1 *Graphical representation of the competitive equilibrium*



The graphic illustration is pretty straightforward for the competitive equilibrium, but it can be shown to hold for any market with little or no price dispersion. We also note that the problem with the measurement of economies of scale lie in the endogenous nature of output, rather than in the specification of the translog cost function. This implies that our conclusion is likely to hold for other specifications as well, including parametric and non-parametric frontier functions. Frontier functions take efficiency differences between firms into account, but they too ignore the fact that these differences have already affected firm behaviour and therefore bias the observed data points. Our

[2] Cost differences are exaggerated for illustrational purposes.

findings are qualitatively consistent with Burnside's (1996) observation that cross-industry restrictions on the parameters induce an upward bias in the estimated degree of returns to scale.

4. What empirical cost functions really tell us

We have shown in the previous section that the implicit assumption that individual output does not depend on individual costs in empirical costs functions leads to incorrect interpretations of the parameters of the cost function. But what *do* these parameters tell us? Let us explore the model from the previous paragraph a little further to find out. Note that the incorrect observation of constant returns to scale stems from the fact that individual firms adjust their output based on cost differences. We rewrite equation (7) to find:

$$\gamma_t = P(X_t)\frac{X_t}{C_t} + X_t\frac{\partial P}{\partial X_t}\frac{X_t}{C_t} - \theta lnX_t \tag{9}$$

And substitute this result into the cost equation:

$$lnC_t = F + P(X_t)\frac{X_t}{C_t}lnX_t + \frac{\partial P}{\partial X_t}\frac{X_t}{C_t}X_t lnX_t - \frac{1}{2}\theta lnX_t^2 \tag{10}$$

For the competitive case, equation (10) reduces to a form that looks somewhat like the translog cost function in equation (2), but has a different interpretation. Note that $P(X_t)\frac{X_t}{C_t}$ boils down to marginal costs divided by average costs, which equals the scale elasticity (see Goh and Yong, 2006, page 849, as well as equation (7) above). Note however that equation (10) suggests that $\frac{X_t}{C_t}lnX_t$ should be in the equation, rather than just lnX_t. This removes the information on returns to scale from the equation, although it can still be evaluated through computation. Furthermore, the estimated parameter for lnX_t^2 equals $-\frac{1}{2}\theta$, which equals the parameter from cost equation (2), but with an opposite sign.

In the non-competitive case, the parameter for $X_t lnX_t$ reflects a scaled version of the (average) conjectural variations parameter. A significant (negative) value for this parameter should normally coincide with positive price-cost margins. Estimated values for this parameter may be compared with estimates from studies aimed at assessing the conjectural variations parameter (e.g. Brander and Zhang (1990) and Oum, Zhang and Zhang (1993)).

5. An empirical illustration

We illustrate out theoretical findings by means of a simple empirical illustration. It is the purpose of this section to give a rough idea of how our proposed adjustment of the translog model works out in practice. We use data on US airfreight-carriers for 2007. Other than passenger transport by air, air freighters deliver a more or less homogenous product. The Bureau of Transport Statistics provides a large amount of data on US air carriers, including financial data and data on outputs.3 We use tables T2 (U.S. Air Carrier TRAFFIC And Capacity Statistics by Service Class) and P-12 (quarterly profit and loss statements). We aggregate the traffic data to quarterly data by carrier and region (BTS distinguishes between Atlantic, Domestic, International, Latin America, Pacific and System) and select carriers that do not transport passengers.4 This leaves us with a dataset of 135 observations, which is fairly small, but large enough for our purpose.

We use operational quarterly expenses as a measure for costs and available ton miles as a measure for output. Table 2 provides the descriptive statistics for our dataset. It is clear from the figures that they cover a wide range of costs and outputs and that both are spread uneven over the population. About 70 percent of the observations lie in the first 10 percentile of the data.

Table 2 *Descriptive statistics*

	Mean	st Dev	Min	Max
Operational expenses (x $1 000)	237 703	656 735	0.03	3 844 896
Available ton miles (millions)	993	1 820	0.45	10 300
number of observations: 135				

We standardized all variables to their own means and estimated equations (2) and (10) from the model in the previous section. Table 3 below provides the estimation results for both models.

Table 3 *Estimation results for logarithm of operational expenses*

variable	parameter	model 1 (equation 2)	model 2 (equation 10)
constant	F	0.212	-8.48 10-4
		(0.3)	(-0.0)
$\ln X$	γ	-0.356	
		(-0.2)	
$\ln X^2$	$\frac{1}{2}\theta$ / - $\frac{1}{2}\theta$	1.095	0.986
		(1.5)	(25.1)
$\frac{X_1}{C_1}\ln X_1$	P		0.004
			(3.9)
$\frac{X_1}{C_1}X_1\ln X_1$	$\frac{\delta P}{\delta X_1}$		-0.034
			(-6.5)
adj R2		0.69	0.89
Log Likelihood		102.9	171.1

[3] See http://www.transtats.bts.gov/databases.asp?Mode_ID=1&Mode_Desc=Aviation&Subject_ID2=0

[4] Carriers that transport passengers, also transport freight as a secundary output. The joint production process –although interesting for a sector study- complicates the analysis and would not serve the illustrational purpose of our analysis. Also note that prices of passenger air services are far from homogenous, whereas our theoretical analysis is based on that assumption.

F	151	351

t-values in brackets

Model 2 clearly performs better from a statistical point of view. None of the parameters in model 1 is significantly different from zero, which is probably due to data quality and the fairly simple nature of our analysis. Model 1 suggests that the industry is operating at increasing returns to scale, but given its poor statistical performance, this is not a very solid conclusion. It does however confirm our earlier statement on the discrepancy between economies of scale in empirical studies and economic theory.

The model parameters of model 2 have the expected sign, except for the parameter of lnX^2, which is expected to have the opposite sign from the same parameter in model 1. On top of that, concavity of the cost function requires this parameter to be negative, which it is not. The other parameters have the expected sign and order of magnitude. The model suggests that the market for air freight transport is not far from being competitive, with $\frac{\delta P}{\delta X_1}$ being close to 0, though the difference is significant. We also find that –on average- $P\frac{X_1}{C_1}$ is close to one, which is consistent with the low value for $\frac{\delta P}{\delta X_1}$.

6. Conclusions

Empirical cost functions do not take into account that cost differences between firms lead to behavioral responses. This omission may lead to erroneous conclusions on economies of scale. Efficient firms are likely to produce more output in markets with homogeneous prices. We show that this behavior may lead the empirical researcher to conclude that the market exhibits increasing returns to scale, whereas each firm in the market actually operates at decreasing returns to scale.

We propose an alternative specification that takes into account behavioral responses and illustrate it using data on air freight carriers. Although the empirical results leave room for improvement, we show that the alternative specification sticks closer to economic theory and provides a better fit.

References

Altunbaş, Y., E.P.M. Gardener, P. Molyneux and B. Moore, 2001, Efficiency in European Banking, *European Economic Review*, **45**, 1931-55.

Altunbaş, Y., P. Molyneux and J. Thornton, 1997, Big-Bank Mergers in Europe: An Analysis of the Cost implications, *Economica* **64**, 317-29.

Berger, A.N., J.H. Leusner and J.J. Mingo, 1997, The Efficiency of bank branches, *Journal of Monetary Economics*, **40**, 141-62.

Brander, J.A. and A. Zhang, 1990, Market Conduct in the Airline Industry: an Empirical Investigation, *RAND Journal of Economics*, **21(4)**, 567-83

Bresnahan, T.F., 1981, Duopoly Models with Consistent Conjectures, *American Economic Review*, **71**, 934-45.

Bresnahan, T.F. and P.C. Reiss, 1991, Entry and Competition in Concentrated Markets, *The Journal of Political Economy*, **99(5)**, 977-1009

Burnside, C., 1996, Production Function Regressions, Returns to Scale, and Externalities, *Journal of Monetary Economics*, **37**, 177-201

Chua, C.L., H. Kew and J. Yong, 2005, Airline Code-share Alliances and Costs: imposing Concavity on Translog Cost Function Estimation, *Review of Industrial Organization*, **26**, 461-87.

Fageda, X., 2006, Measuring Conduct and Cost Parameters in the Spanish Airline Market, *Review of Industrial Organization*, **28**, 379-99.

Farrel, J. and C. Shapiro, 2000, Scale Economies and Synergies in Horizontal Merger Analysis, Competition Policy Center Working Paper CPC00-15.

Fries, S. and A. Taci, 2005, Cost Efficiency of Banks in Transition: Evidence from 289 banks in 15 post-communist countries, *Journal of Banking and Finance*, **29**, 55-81.

Goh, M. and J. Yong, 2006, Impacts of Code-share alliances on Airline Cost Structure: A Truncated third-order Translog Estimation, *International Journal of Industrial Organization*, **24**, 835-66.

Hansen, K.K. and J. Zwanziger, 1996, Marginal Costs in General Acute Care Hospitals: A Comparison among California, New York and Canada, *Health Economics*, **5**, 195-216.

Hughes, J.P. and L.J. Mester, 1993, A Quality and Risk Adjusted Cost Function For Banks: Evidence on the “Too-Big-to-Fail” Doctrine, *The Journal of Productivity Analysis*, **4**, 293-315

Hughes, J.P. and L.J. Mester, 1998, Bank Capitalization and Cost: Evidence of Scale Economies in Risk Management and Signalling, *The Review of Economics and Statistics*, 314-25

Humphrey, D.B., 1993, Costs and Technical Change: Effects from Bank Deregulation, *The Journal of Productivity Analysis*, **4**, 9-34.

Malhotra, D.K. K., V.B. Marisetty and M. Ariff, 2001,Economics of Scale in the Retail Superannuation Funds in Australia. EFMA 2002 London Meetings.

McCarthy, P. and A. Urmanbetova, 2006, Production and Cost in the Pulp and Paper Industry: A Translog Cost Function Analysis, CPBIS Working Paper.

Mountain, D.C. and H. Thomas, 1999, Factor Price Misspecification in Bank Cost Function Estimation, *Journal of International Financial Markets, Institutions & Money*, **9**, 163-82.

Oum, T.H., A. Zhang and Y. Zhang, 1993, Inter-Firm Rivalry and Firm-Specific Price Elasticities in Deregulated Airline Markets. *Journal of Transport Economics and Policy*, **May 1993**, 171-92

Panzar, J.C. and R.D. Willig, 1977, Economies of Scale in Multi-output Production, *Quarterly Journal of Economics,* **91(3),** 481-93

Pels, E. P. Nijkamp and P. Rietveld, 2003, Inefficiencies and Scale Economies of European Airport Operations, *Transportation Research Part E*, **39**, 341-61

Perry, M.K., 1982, Oligopoly and Consistent Conjectural Variations, *Bell Journal of Economics*, **14**, 301-14

Siegfried, J.J. and L.B. Evans, 1994, Empirical Studies of Entry and Exit: A Survey of the Evidence, *Review of Industrial Organization*, **9**, 121-55

Torres, M. and C.J. Morrison Paul, 2006, Driving forces for consolidation or fragmentation of the US water utility industry: A Cost Function Approach with Endogenous Output, *Journal of Urban Economics,* **59**, 104-20.

Varian, H., 2006, *Intermediate Microeconomics*, 7nd edition, Norton and Company, New York.